AFRL-RI-RS-TR-2011-238

GENERATING REALISTIC ENVIRONMENTS FOR CYBER OPERATIONS DEVELOPMENT, TESTING, AND TRAINING

DARTMOUTH COLLEGE

OCTOBER 2011

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2011-238 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/ /s/

JOHN PERRETTA
Work Unit Manager

WARREN H. DEBANY, JR., Technical Advisor
Information Exploitation and Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
October 2011	Final Technical Report	May 2010 – May 2011

4. TITLE AND SUBTITLE

GENERATING REALISTIC ENVIRONMENTS FOR CYBER OPERATIONS DEVELOPMENT, TESTING, AND TRAINING

5a. CONTRACT NUMBER
FA8750-10-1-0039

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
62303E

6. AUTHOR(S)

Vincent H. Berk
Ian Gregorio-de Souza
John P. Murphy

5d. PROJECT NUMBER
NTGS

5e. TASK NUMBER
00

5f. WORK UNIT NUMBER
10

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dartmouth College/Thayer School of Engineering
8000 Cummings Hall
Hanover NH 03766

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/Information Directorate
Rome Research Site/RIGB
525 Brooks Road
Rome NY 13441

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TR-2011-238

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research was a focused effort to quantify what measurements and metrics can be used to evaluate and improve traffic generator realism. Due to the complexity of network traffic the number of possible measurements are endless. To cope with this complexity we grouped over 150 different measurements into three broad categories based on complexity and required realism for a specific task. This allowed us to measure quantities in a sampling of both real network traffic, and a representative sampling of generated traffic, in a structured manner.

15. SUBJECT TERMS
Realistic Simulated Network, Traffic Generation Analysis, Characterization Metrics, Role-Based

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			JOHN PERRETTA
U	U	U	UU	20	19b. TELEPHONE NUMBER (*Include area code*) N/A

Table of Contents

List of Figures

List of Tables

1.0 SUMMARY

Effective cyber operations research depends in large part on the ability to reliably reproduce test conditions, such that methodologies and technologies may be evaluated for effectiveness and then subsequently retuned and re-evaluated for improvement. The required availability of a reproducible test environment often leads both researchers as well as operators to resort to using simulators for network traffic and host activity. Careful examination of the resulting traffic, however, often reveals substantial shortcomings in realism. Some differences between simulated and real traffic data are visually illustrated in Figure 1. New (red) and concurrent (blue) sessions per second for a two day time period on a real network (bottom), vs. a LARIAT simulated network (top) are plotted in the figures. These visual depictions illustrate that simulated traffic, even when carefully crafted, can fail to be similar to real traffic even in superficial ways. Moreover, quantitative metrics are needed to capture differences for scalability, objectivity and extensibility.

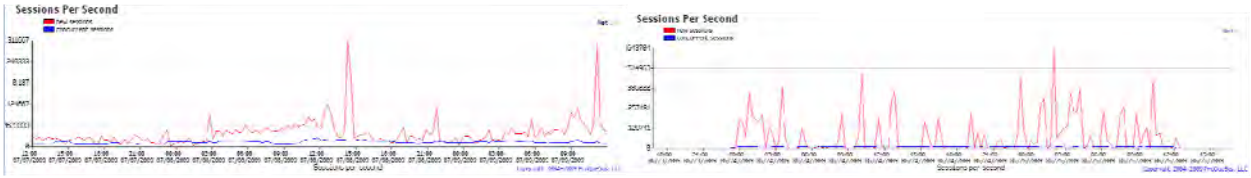


Figure 1. Concurrent Session Per Second in Real Versus Simulated Traffic

This research, supported by DARPA award FA8750-10-1-0039, was a focused effort to quantify what measurements and metrics can be used to evaluate and improve traffic generator realism. This final report outlines the findings and recommendations from the yearlong effort, and the industry workshop that was held on November 7, of 2010. Due to the complexity of network traffic the number of possible measurements are endless. To cope with this complexity we grouped over 150 different measurements into three broad categories based on complexity and required realism for a specific task (Table 1). This allowed us to measure quantities in a sampling of both real network traffic, and a representative sampling of generated traffic, in a structured manner. Our findings were presented at a two-day industry workshop where feedback was gathered. This feedback was used to direct the final 6 months of the project, and ultimately shaped our conclusions as presented in this final report. To augment our findings, the reference implementation of the various algorithms that were used in this research were delivered in a VMWare virtual machine. A browser-friendly user interface allows the user to explore similarities and realism of generated datasets compared to real-world captured data. This system image was delivered separately to DARPA on Friday April 8th, 2011. *(Detailed findings from the workshop were reported in a separate document earlier in the performance period.)*

Table 1. Three Broad Categories of Traffic Simulation Realism Requirements

Task (examples)	Realism Level
Infrastructure Load Testing: Routers, switches Server capacity DOS training, experimentation Botnet deployment	*Statistical Realism:* Observe the ”power-law” IP address, port range distributions “Lorem ipsum” content Services do not respond realistically
Offensive Operations Training: Exploit development, testing Coarse infiltration, exfiltration Privilege escalation experiments (Coarse: the value of information is not considered)	*Content Realism:* Email through SMTP, with real addresses Websites contain HTML with working links Files on shares contain random data Servers run actual services OS realism required
Intelligence Analyst Training: Information gathering (espionage) Business disruption (sabotage) Insider threat detection (Fine: value of information is key differentiator)	*Behavioral Realism:* Social networks simulated Users have tasks, objectives, intent Files contain realistic information Value of information is relevant and different for all business processes

2.0 INTRODUCTION

There is a range of applications that benefit from realistically generated network traffic: from demonstrating the reliability of network hardware to testing new network analysis software to providing a background of traffic for experiments. Ultimately, the most realistic traffic is captured on a live network with the right scale and architecture, the right users behaving naturally and performing the expected tasks. Institutional, physical, and legal concerns cause such traffic captures to be generally unavailable. This leaves us to decide subjectively what is *sufficiently realistic* for a given task. Fortunately, not all aspects of realism are germane to all applications. There are many aspects to network traffic, any one of which may be relevant for traffic to be considered *sufficiently realistic* for a specific task, and any two applications are unlikely to share the same realism requirements, generally aiming toward realism in terms of a specific OSI Layer or Layers (Figure 2). Examples of realism measurements include inter-packet timing and drop rate, or traffic volume by application, or which web sites individual users visit or realistic social networks of users who exchange emails, instant messages, and web links according to task and hierarchy.

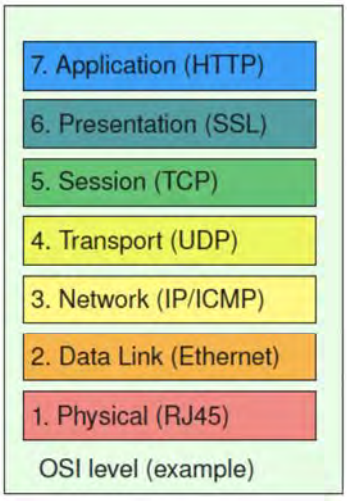


Figure 2. OSI Model

There are many different kinds of traffic generation software available, both free and commercially. Hardware-based solutions are available for longer-term traffic generation, such as BreakingPoint, and network-on-chip generators[1]. Other traffic generators, such as Swing[2] or Harpoon[3] use automated or semi-automated network collectors, and tune a set of internal rules to generate traffic that matches observed statistics. These tend to focus on lower-level aspects of the OSI Layers, such as traffic volumes, packet timings, drop rates, etc. They are frequently referred to as "network emulators" and some of them (such as Swing) are not intended to be used alone, but instead to generate background traffic only.

A number of generators have been proposed specifically for applications related to network security and testing network security software. These applications often focus on user behavior. In a real network, much of the behavior is emergent in nature, the result of many users acting individually and in concert, and network simulations represent an attempt to recreate that emergent behavior by simulating an agent-based environment on a single computer or a network of semi-automated systems. Among them are the Lincoln Adaptable Real-time Information Assurance Testbed (LARIAT)[4], Skaion Corporation's Traffic Generation System (TGS), and the LLSIM network simulator[5].

DARPA's National Cyber Range has introduced a powerful new experimental facility for evaluating cyber security technologies. Traffic generation aimed specifically at this purpose is a significant part of this effort. In order to quantify and evaluate candidate technologies, it is necessary to adapt existing methods of traffic analysis and develop new ones to determine how realistic the traffic output is, and how different traffic generators are from each other and from real traffic conforming to desired specifications.

In order to evaluate traffic generators and simulators, it is necessary to find methods to analyze each to quantify the differences in output to real traffic. A number of methods are used to analyze traffic for other purposes, which suggest approaches to this problem. The Swing traffic emulator is based on sophisticated statistical analysis on low-level network characteristics to determine how best to reproduce them. Techniques for intrusion detection, such as n-gram analysis[6] focus on packet payload, and can potentially be repurposed for use at higher OSI

levels to examine the realism of application traffic. Behavior-based insider threat detection techniques for profiling users by work hours[7] or social groups[8] that are ordinarily used to determine when users are acting strangely with respect to their histories can be repurposed to examine work hour normality for a population of simulated agents.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

To offer a guideline for the level of realism required for a specific task, we have divided the measurements and metrics into three broad categories (shown in Table 1). Each category captures a deeper, and more complex level of realism, and reaching satisfactory levels of realism in each subsequent category therefore gets increasingly more difficult to accomplish. We consider the following distinction: measurements are quantities produced by sensors while metrics are quantitative comparisons between measurements.

We continue by summarizing a number of representative measurements and metrics for each of the three broad categories. We demonstrate their typical values measured on a large college campus, as well as a small research lab. We also applied our measurements to a number of well-established traffic generators to show the similarities and differences.

3.1 Statistical Realism

The measurements in this category focus solely on the first 6 layers of the OSI model. Specifically, these statistical properties can be derived from network traffic up to, and including, the TCP/UDP/ICMP/etc protocol headers, without deeper inspection of the transmitted application payload. Examples include the statistical properties of network connections, addressing, and port allocation. For instance, the typical connection pattern, connection volume by time of day, balance between initiation and reception of connections, and unique destination count will differ greatly between an email server, a web server, and a user laptop computer. Also, on a host-by-host comparison we would expect to see repeating connection patterns, in terms of time of day, volume (packets and bytes) and protocols used (for instance 2 email servers communicating).

3.2 Content Realism

The measurements in this category focus on the 7th layer of the OSI model: application content. For instance, a user will frequent a small number of websites very regularly, often at predictable times of day, and a larger number of sites very infrequently. Similarly, the email network, and an instant messenger social network will have a distribution of frequently communicating addresses, and a larger number of less frequently contacted addresses. In order to determine the properties of this class, one must inspect the payloads of the TCP/UDP/ICMP/etc sessions to analyze protocol content.

3.3 Behavioral Realism

The measurements in this category are defined in terms of the first two categories, except the bar has been raised from straightforward statistical analysis to cause and effect analysis, where statistical measurements are conditioned upon other measurements and states. The mathematical techniques include Markovian models, time conditioned sequences, etc. These patterns are more complex, in that each Internet user has a unique pattern of websites, email servers, and work times, etc. which identifies the user. For instance, it is not realistic, within the context of most

real networks, for all users on a simulated network to visit the same websites at approximately the same time of day. This is further complicated by the fact that network traffic follows specific patterns of behavior that are typical to a given person at a particular time in their personal work processes.

For instance, towards the end of the fiscal year, many people are moving budgets around, doing related searches, and are contacting people that they otherwise will not have a reason to talk to.

Additionally, this category also includes a subgroup that we refer to as *contextual realism*. In addition to each user possessing a unique signature based on their usage of network resources over time, each network resource has contextual characteristics that imply a relation among the entities that use it. For example, frequent usage of a printer is often indicative of location, as frequent usage of certain network shares is indicative of membership in a work group.

Furthermore, individual signatures may depend strongly on group dynamics over time; e.g. time of day patterns for certain groups of individuals show strong correlations. For example, students have class in the same location and time each week.

4.0 RESULTS AND DISCUSSION

In order to illustrate the significance of the three categories of measurements, we implemented a representative subset of metrics from each category and applied them to sample network traffic datasets. This section describes the relevant metrics and measurements used and shows some results from our experiments.

4.1 Statistical and Content

It has been established through published works[9], that both natural and man-made occurring activities follow the power-law. Common cyber observables in particular have been shown to strongly follow the power-law distributions[10][11]. Typical generative values for these power-law distributions vary from network to network, and thus measuring the shape of the curve carries more weight than the typical values. Our research of numerous computer networks has generally shown the existence of these power-laws in almost all measurements taken, however, the generative values may vary a great deal between different networks. For the statistical and content categories we made use of the methods laid out by Clauset, Shalizi et.al. in their empirical study of the Power Law distribution to quantify similarity in traffic based on the following statistical measurements for each of 15 different traffic groups (HTTP, mail, fileshare, etc):

- **Sessions per host:** The distribution of sessions initiated by a host over time. Some hosts will initiate more web sessions than others.
- **Bytes per session:** The average size of a session is distributed over all hosts.
- **Packets per session:** Similarly, the average number of packets for a session will vary from host to host.
- **Unique peers per host:** The number of unique servers contacted per client.

and the following content measurements:

- **Unique URLs visited:** The distribution of unique URLs visited by hosts.
- **Distribution of URIs visited:** Each host will visit certain URIs with specific URNs other than the root of the website.
- **Emails Sent by user:** Users will send a varying amount of emails to other users.
- **Unique Email Subjects:** Email threads will either die out quickly after a couple of messages or linger as messages are replied to and forwarded.
- **FTP and SMB Unique Filenames:** The frequency with which files are accessed and/or modified on fileshares available to multiple people. Some files will be accessed rarely, by few people, while others will be accessed at a high frequency by a large number of users.

For this article, we present statistical results based on web (HTTP) and mail (SMTP) traffic from two live traffic captures, and one generated traffic capture for comparison. Content results are based on the Enron email dataset [12], and a simulated SMTP traffic.

Clauset et.al showed that power-law distributions may be described by the basic equation:

$$p(x) = Cx^{-\alpha} \tag{1}$$

where α is a scaling constant and C is a normalization constant. Since Equation 1 diverges as $x \to 0$, another parameter, $x_{\min} > 0$, which factors into the calculation of the normalization constant was used to specify the minimum range above which the power-law was valid. Figure 3 shows plots of power-law fits for a set of the statistical measurements across the three capture types. Visual inspection, confirms the general shape of the power-law distribution in each of the measurements. As expected, the generative values describing the distribution – α and $x_{\min}$ – differ by environment.

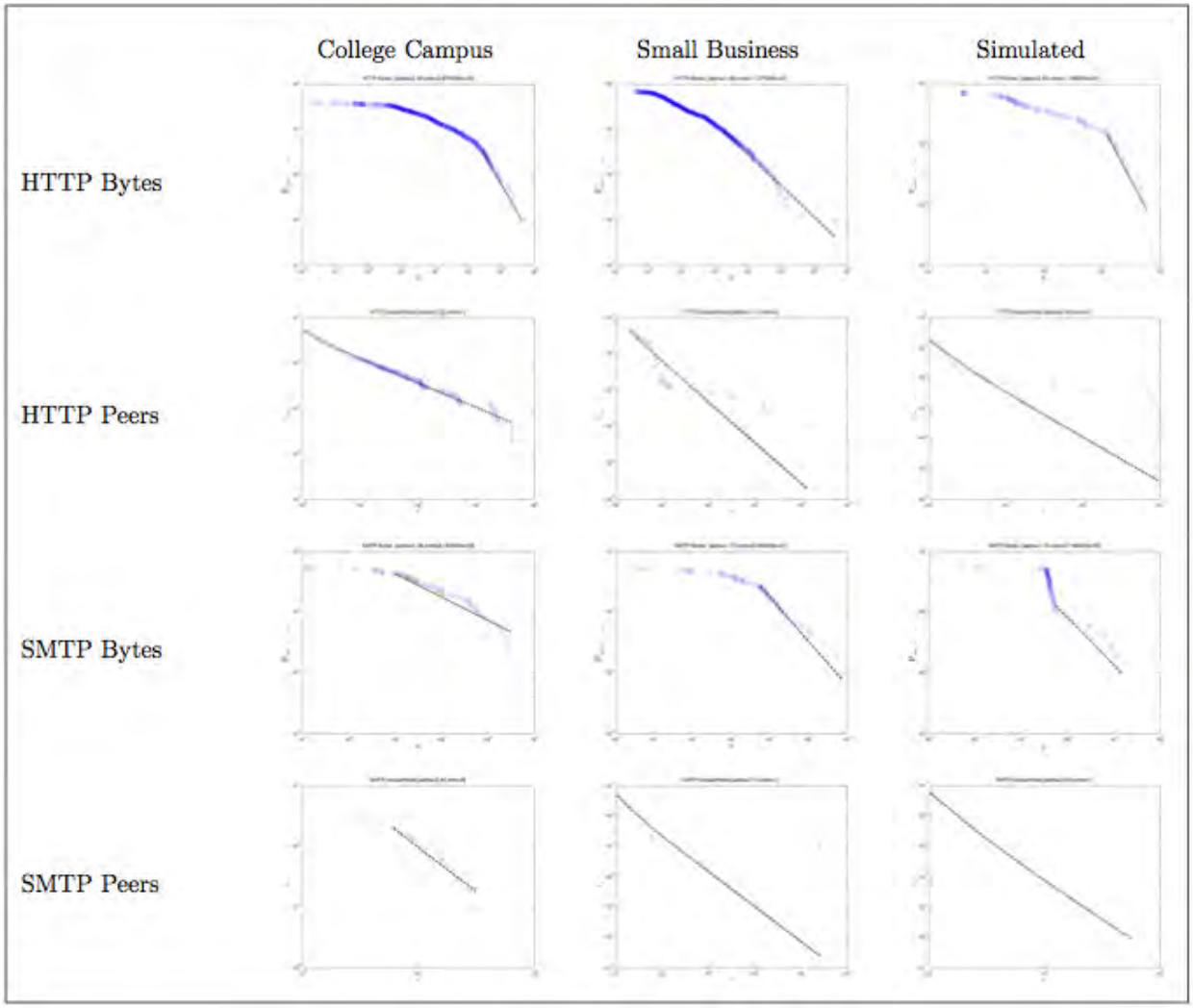


Figure 3. Power-law Fits for Statistical Measurements from Three Network Traffic Types

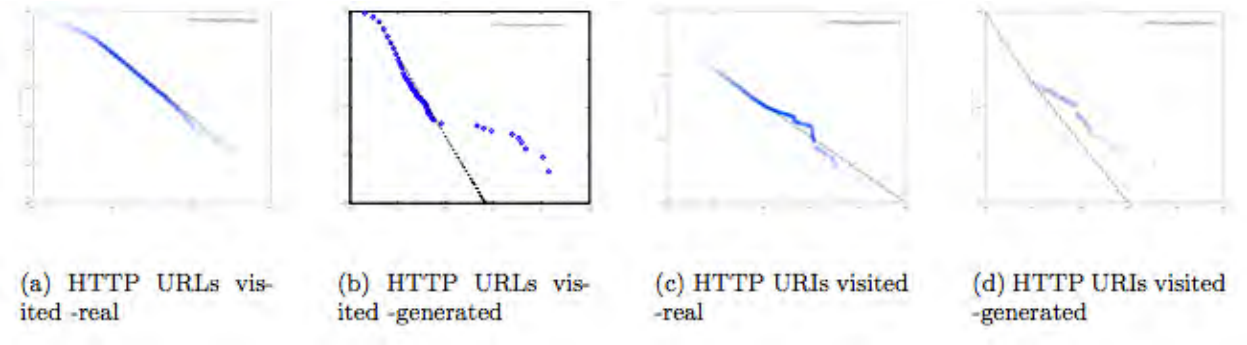
Another important statistical measurement is one that shows the statistical dispersion in the created traffic. A measure of statistical dispersion is zero if all data is identical and increases as the data becomes diverse. On a large network, the diversity of user activities should be evident in the network traffic and conveyed in the five measurements mentioned previously. A common measure of statistical dispersion is the standard deviation. However, to facilitate better comparisons with measurements having different units (bytes, packets, seconds, etc), a dimensionless measure - the coefficient of variation was be used. The coefficient of variation is defined as the ratio of the standard deviation to the mean[1].

$$C_v = \frac{\sigma}{|\mu|} \tag{2}$$

Deeper inspection of the traffic shows that although the analyzed generated traffic conforms sufficiently to the power-law distribution, we begin to notice a lack of realism when we examine

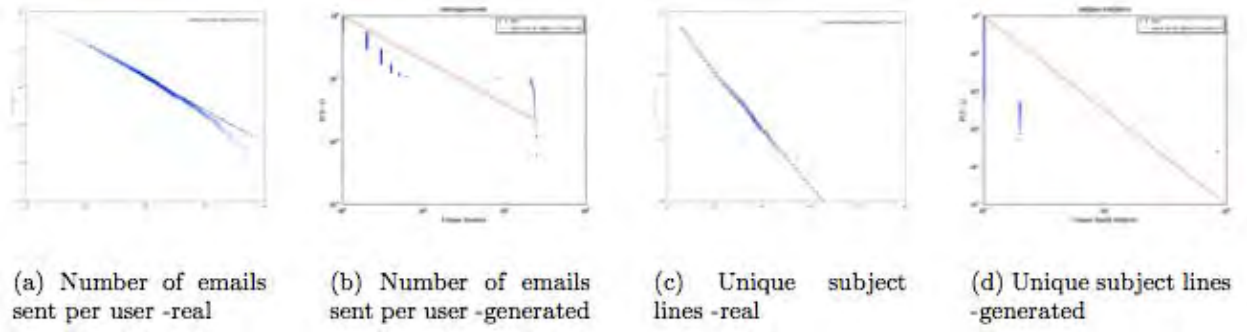
[1] http://en.wikipedia.org/wiki/Coefficient_of_variation

the content of the network packets. The power-law fits for HTTP URLs (hosts) visited and HTTP URIs (Figure 4) show larger deviations from the fit lines than were observed in the statistical measurement fits. The same results can be observed when comparing emails sent by users and unique subject line of an email corpus as shown in Figure 5. These deviations begin to reveal the lack of realism found in network traffic generators.



(a) HTTP URLs visited -real (b) HTTP URLs visited -generated (c) HTTP URIs visited -real (d) HTTP URIs visited -generated

Figure 4. HTTP Content Analysis: Real Versus Generated URLs and URIs



(a) Number of emails sent per user -real (b) Number of emails sent per user -generated (c) Unique subject lines -real (d) Unique subject lines -generated

Figure 5. SMTP Content Analysis: Real (Enron) Versus Generated Emails Sent and Unique Subject Lines

4.2 Behavioral

Measuring *Behavioral Realism* is a subjective task, especially as it pertains to the value of the measured quantities. Since it is impossible to capture all cause-and-effect relationships based on the user's objectives on the network, we have settled on a number of different measurements indicated by collective agreement to be valuable during the industry workshop. We measure the first dimension of behavior by comparing graph-based social network analysis metrics and temporal activity of traffic captures. We divided traffic for social network metrics into categories similar to the categories used for Statistical Measures. Traffic from each category was then used to build a directed graph on which the following metrics are computed:

- **Degree Centrality:** The degree centrality of a node is the number of direct connections the node has. A high degree centrality indicates a very active node.
- **Closeness Centrality:** Closeness centrality measures how easily a node can access other nodes in the network. A high closeness value indicates short path to other nodes in the network and high visibility of network activity.

- **Betweenness Centrality:** Betweenness centrality measures the extent to which a node acts as a bridge between other nodes on the network. Nodes with high betweenness values usually act as bridges between clusters.

In Figure 6, we show three such graphs for HTTP traffic. These illustrate differences that will be captured in the above metrics. For example, the degree centrality in Figure 6(a) and Figure 6(b) would be much higher than Figure 6(c). Figure 6(c) is however going to have a higher betweenness score since we can observe nodes acting as bridges between clusters. These bridge nodes are visibly absent from the other graphs. The social network layout of traffic will differ from environment to environment and therefore in determining how realistic a capture is, there needs to some ground truth comparison. High or low scores in these categories do not necessarily indicate real or unrealistic traffic, but in comparison to the environment being mimicked can indicate how much realism has been captured from the real environment.

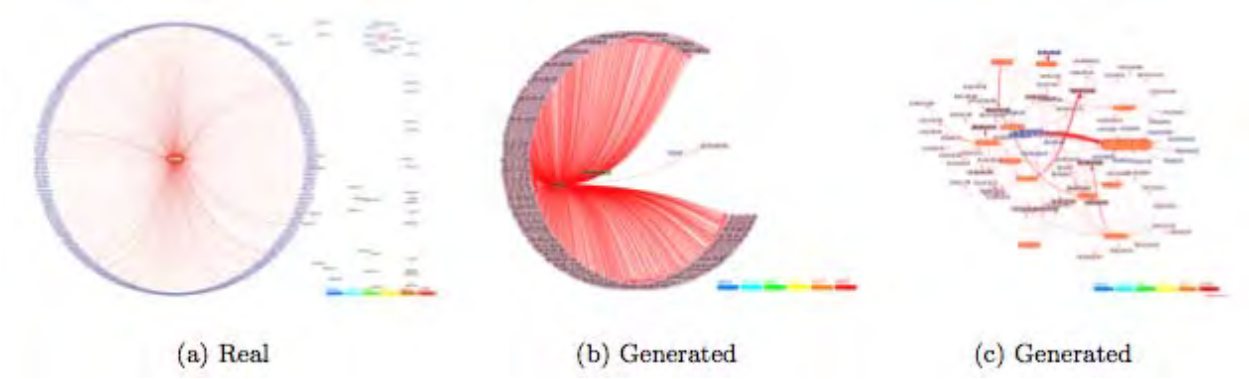


Figure 6. HTTP Network Graphs for Three Different Traffic Captures

Our second dimension of behavioral measurements is *Temporal Activity* measured by comparing the hourly probability distributions of traffic volume per user (total bytes per hour). A histogram of the traffic volume distribution gives a snapshot of when the network is being actively utilized. In a realistic environment, we expect to see peaks during period of high human activity and lower volume levels when the users of the network are away (gone home/asleep, weekends). It is not abnormal to still observe moderate levels of activity outside “work hours” since most servers are still active and background jobs (backups, updates, etc) are usually scheduled for slow hours. However, a pretty flat distribution of traffic volume over the course of a day is highly unlikely to be real. Figure 7 shows a sample temporal profile from real and generated network captures. The difference is highly visible.

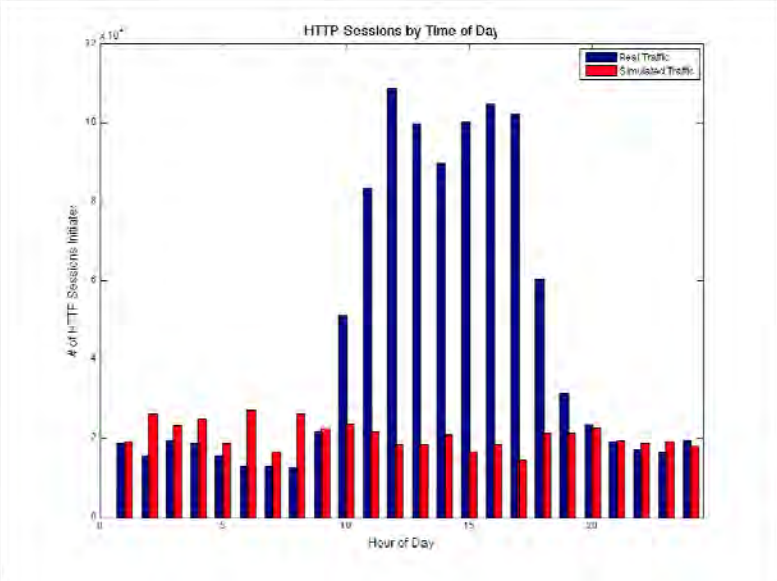


Figure 7. Temporal Activity for Real and Generated Traffic

4.3 The Traffic Capture Comparator

As part of this research effort a large number of algorithms were implemented to compute and collect the results for the recommendations included in this report. Although a software delivery was not a deliverable of this contract, we believe the sponsor may benefit from the code that was produced. We collected and organized the code in a virtual machine with a browser-friendly interface, capable of comparing traffic captures for similarity on a large range of metrics from the statistical, content, and behavioral realism categories. The user can upload traffic captures from a real-world network, and compare them against traffic captures from a generated dataset to assess realism in many aspects. (*Note: The VMWare virtual machine was delivered to DARPA on Friday April 8, 2011).*

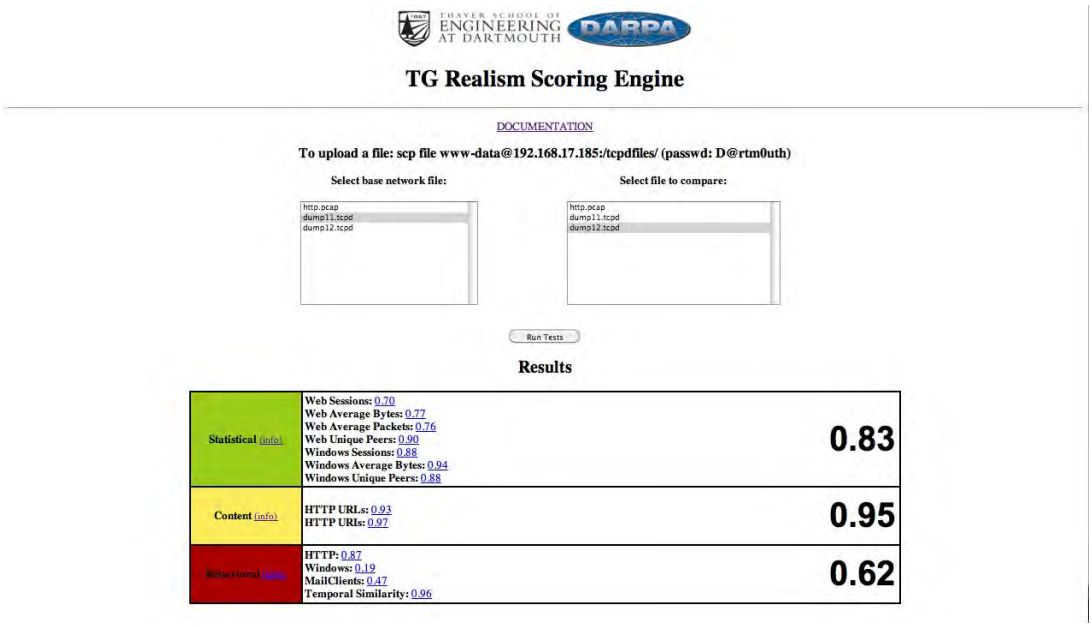


Figure 8. User interface of traffic capture comparator

The statistical measurements comprise the largest number of single measurements in the scoring engine. Realism is judged on a variety of distributions for a collection of different application protocols on the network. Specifically, the distributions in the *base* network are compared against distributions in the *test* network to decide similarity. For each of the distributions below, the Coefficient of Variation is computed as a measure of dispersion present among the hosts:

- **Web sessions per host:** The distribution of web sessions initiated by a host over time. Some hosts will initiate more web sessions than others.
- **Bytes per web session:** The average size of a web session is distributed over all hosts according to a powerlaw.
- **Packets per web session:** Similarly, the average number of packets for a web session will vary from host to host.
- **Unique web peers per host:** The number of unique servers contacted per client varies according to a powerlaw distribution.
- **SMTP sessions per host:** same as web.
- **Average bytes per SMTP session:** same as web.
- **Average packets per SMTP session:** same as web.
- **Unique SMTP peers per host:** same as web.
- **Windows sessions per host:** same as web. Windows protocols include ports 135-139.
- **Average bytes per Windows session:** same as web.
- **Average packets per Windows session:** same as web.
- **Unique Windows peers per host:** same as web.

Content realism is measured by comparing distributions of specific traffic content between the *base* and the *test* traffic captures. Included measurements:

- **Unique URLs visited:** The distribution of unique URLs visited by hosts in the capture. This distribution should follow the powerlaw.
- **Distribution of URIs visited:** Each host will visit certain URIs with specific URNs other than the root of the website. The distribution of visiting frequency per URI is powerlaw distributed.
- **Emails Sent by user:** Users will send a varying amount of emails to other users. The distribution of the number of emails sent by each user should follow the powerlaw.
- **Unique Email Subjects:** Email threads will either die out quickly after a couple of messages or linger as messages are replied to and forwarded. The distribution of unique subject lines is powerlaw distributed.
- **FTP and SMB Unique Filenames:** The frequency with which files are accessed and/or modified on fileshares available to multiple people should be powerlaw distributed. Some files will be accessed rarely by few people while others will be accessed at a high frequency by a large number of users.

Behavioral realism is measured by comparing graph-based social network analysis metrics and temporal activity between the *base* and the *test* traffic captures. Traffic for social network metrics is divided into categories similar to the categories used for Statistical Measures. Traffic from each category is used to build a directed graph on which the following metrics are computed:

- **Degree Centrality:** The degree centrality of a node is the number of direct connections the node has. A high degree centrality indicates a very active node. The average degree centrality of nodes is compared
- **Closeness Centrality:** Closeness centrality measures how easily a node can access other nodes in the network. A high closeness value indicates short path to other nodes in the network and high visibility of network activity. The average closeness of nodes is compared.
- **Betweenness Centrality:** Betweenness centrality measures the extent to which a node acts as a bridge between other nodes on the network. Nodes with high betweenness values usually act as bridges between clusters. The average betweenness of nodes is compared.

Finally, temporal Similarity between the traffic captures is measured by comparing the hourly probability distribution of traffic volume (total bytes per hour).

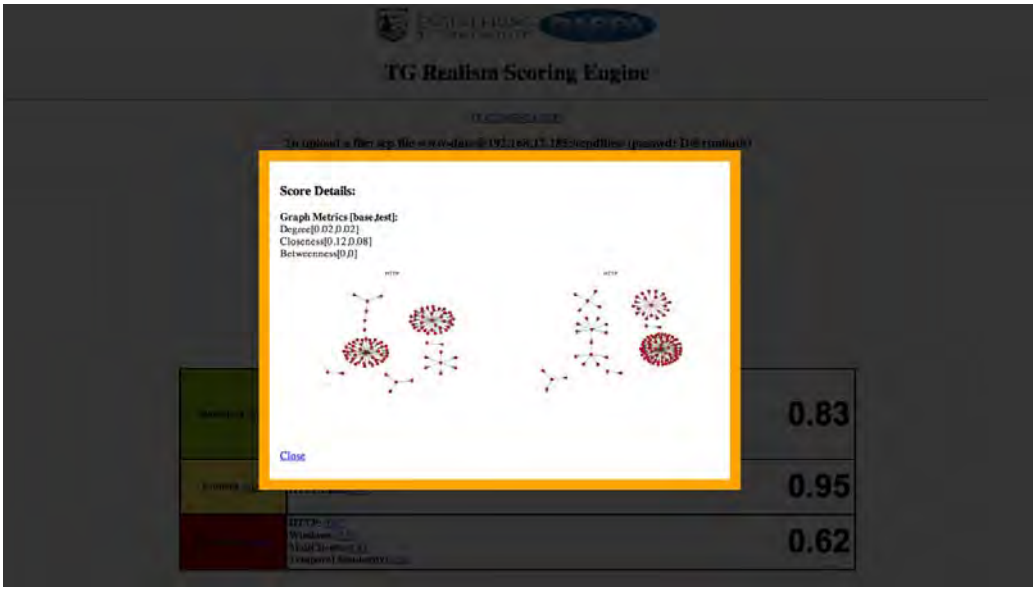


Figure 9. Detailed view of social net behavioral analysis scoring

5.0 CONCLUSIONS AND RECOMMENDATIONS

Truly realistic traffic generation can arguably only be achieved by simulating the complete universe. It is therefore only possible to achieve realism asymptotically for any given purpose. Recent advances in behavioral modeling, however, open the door for *intent-driven* traffic generation, where actors are simulated based on objectives, desires, business processes, and environmental factors. Such traffic simulation will be amply sufficient for the large majority of tasks, up to and including cyber operations training.

Ultimately, observed network traffic is a consequence of human actors trying to achieve an objective. Our objectives are a direct consequence of our environment; people we need to communicate with to achieve our goals, relationships we develop, messages we send, and documents we store, retrieve and modify. By describing business processes, and environmental factors in a behavioral modeling framework, we can instantiate actors that trigger network traffic which is an order of magnitude more complex and realistic than current traffic generators are capable of.

Our work in measurements and metrics of traffic generators has only explored the fringes of what is possible in realistic traffic generation with regard to training cyber operatives, and developing new cyber warfare strategies. Although many experts have weighed in on the decisions of what to measure, we do not claim our approach is complete and exhaustive. Rather we intend this work as a guideline to deciding the level of realism required for a specific task, and the factors to consider when implementing a traffic generator.

6.0 REFERENCES

[1] Mahadevan, S., Angiolini, F., Storgaard, M., Olsen, R. G., Sparso, J., and Madsen, J., "A network traffic generator model for fast network-on-chip simulation," in *Proceedings of the conference on Design, Automa- tion and Test in Europe - Volume 2*, DATE '05, 780–785, IEEE Computer Society, Washington, DC, USA (2005).

[2] Vishwanath, K. and Vahdat, A., "Swing: Realistic and responsive network traffic generation," *Networking, IEEE/ACM Transactions on* 17, 712 –725 (june 2009).

[3] Sommers, J. and Barford, P., "Self-configuring network traffic generation," in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, IMC '04, 68–81, ACM, New York, NY, USA (2004).

[4] Rossey, L. M., Cunningham, R. K., Fried, D. J., Rabek, J. C., Lippmann, R. P., Haines, J. W., and Zissman, M. A., "Lariat: Lincoln adaptable real-time information assurance testbed," in *In IEEE Proc. Aerospace Conference*, 2671–2682 (2001).

[5] Haines, J., Goulet, S., Durst, R., and Champion, T., "Llsim: network simulation for correlation and response testing," in *Information Assurance Workshop, 2003. IEEE Systems, Man and Cybernetics Society*, 243 – 250 (june 2003).

[6] Wang, K. and Stolfo, S., "Anomalous payload-based network intrusion detection," in **Recent Advances in Intrusion Detection**, Jonsson, E., Valdes, A., and Almgren, M., eds., Lecture Notes in Computer Science 3224, 203–222, Springer Berlin / Heidelberg (2004). 10.1007/978-3-540-30143-1 11.

[7] Murphy, J. P., Berk, V. H., and de Souza, I. G., "Effectively identifying user profiles in network and host metrics," 7666(1), 766607 (2010).

[8] de Souza, I. G., Berk, V., and Barsamian, A., "Using principal component analysis for selecting network behavioral anomaly metrics," 7666(1), 766605 (2010).

[9] Clauset, A., Rohilla Shalizi, C., and Newman, M. E. J., "Power-law distributions in empirical data," ArXiv e-prints (June 2007).

[10] Huberman, B. A. and Adamic, L. A., "The nature of markets in the world wide web," *Computing in Economics and Finance* 1999 521, Society for Computational Economics (1999).

[11] Albert, R., Jeong, H., and Barabasi, A.-L., "Internet: Diameter of the world-wide web," nat 401, 130–131 (Sept. 1999).

[12] Shetty, J. and Adibi, J., "The Enron email dataset database schema and brief statistical report," Information Sciences Institute Technical Report, University of Southern California (2004).

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

LARIAT	Lincoln Adaptable Real-time Information Assurance Testbed
DARPA	Defense Advanced Research Projects Agency
DOS	Denial Of Service (type of attack)
SMTP	Simple Mail Transfer Protocol
HTML	HyperText Markup Language
OS	Operating System
OSI	Open Systems Interconnect (OSI model)
SSL	Secure Sockets Layer
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
IP	Internet Protocol
ICMP	Internet Control Message Protocol
TGS	Traffic Generation System (by Skaion Corp.)
NCR	National Cyber Range
URL	Uniform Resource Locator
URI	Uniform Resource Identifier
FTP	File Transfer Protocol
SMB	Server Message Block (Protocol)